

*11*  
*2*

## STIC Biotechnology Systems Branch

### RAW SEQUENCE LISTING ERROR REPORT

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 10/540,615  
Source: PCT  
Date Processed by STIC: 07-11-2005

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.  
PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:  
1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,  
2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY  
FOR CRF SUBMISSION AND PATENTIN SOFTWARE QUESTIONS, PLEASE CONTACT MARK SPENCER, TELEPHONE: 571-272-2510; FAX: 571-273-0221

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 4.2.2 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:

<http://www.uspto.gov/web/offices/pac/checker/chkrnote.htm>

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail. Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom. Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<http://www.uspto.gov/ebc/efs/downloads/documents.htm>), EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450
3. Hand Carry, Federal Express, United Parcel Service, or other delivery service (EFFECTIVE 01/14/05): U.S. Patent and Trademark Office, Mail Stop Sequence, Customer Window, Randolph Building, 401 Dulany Street, Alexandria, VA 22314

Revised 01/24/05

## Raw Sequence Listing Error Summary

<u>ERROR DETECTED</u>	<u>SUGGESTED CORRECTION</u>	<u>SERIAL NUMBER:</u> <u>10/540,615</u>
<b>ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE</b>		
1 <input type="checkbox"/> Wrapped Nucleics <input type="checkbox"/> Wrapped Aminos	The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor <b>after</b> creating it. Please adjust your right margin to .3; this will prevent "wrapping."	
2 <input type="checkbox"/> Invalid Line Length	The rules require that a line <b>not exceed</b> 72 characters in length. This includes white spaces.	
3 <input type="checkbox"/> Misaligned Amino Numbering	The numbering under each 5 <sup>th</sup> amino acid is misaligned. Do <b>not</b> use tab codes between numbers; use <b>space characters</b> , instead.	
4 <input type="checkbox"/> Non-ASCII	The submitted file was <b>not</b> saved in ASCII(DOS) text, as <b>required</b> by the Sequence Rules. Please <b>ensure your subsequent submission is saved in ASCII text</b> .	
5 <input type="checkbox"/> Variable Length	Sequence(s) _____ contain n's or Xaa's representing more than one residue. <b>Per Sequence Rules</b> , <b>each n or Xaa can only represent a single residue</b> . Please present the <b>maximum</b> number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.	
6 <input type="checkbox"/> PatentIn 2.0 "bug"	A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) _____. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. <b>This applies to the mandatory &lt;220&gt;-&lt;223&gt; sections for Artificial or Unknown sequences</b> .	
7 <input type="checkbox"/> Skipped Sequences (OLD RULES)	Sequence(s) _____ missing. If intentional, please insert the following lines for <b>each</b> skipped sequence: (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading) (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) This sequence is intentionally skipped  Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to <b>include</b> the skipped sequences.	
8 <input type="checkbox"/> Skipped Sequences (NEW RULES)	Sequence(s) _____ missing. If <b>intentional</b> , please insert the following lines for <b>each</b> skipped sequence. <210> sequence id number <400> sequence id number 000	
9 <input type="checkbox"/> Use of n's or Xaa's (NEW RULES)	Use of n's and/or Xaa's have been detected in the Sequence Listing. Per 1.823 of Sequence Rules, use of <220>-<223> is <b>MANDATORY</b> if n's or Xaa's are present. In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.	
10 <input type="checkbox"/> Invalid <213> Response	Per 1.823 of Sequence Rules, the <b>only valid</b> <213> responses are: <u>Unknown, Artificial Sequence, or scientific name (Genus/species)</u> . <del>&lt;220&gt;-&lt;223&gt; section is required when &lt;213&gt; response is Unknown or is Artificial Sequence</del>	
11 <input type="checkbox"/> Use of <220>	Sequence(s) _____ missing the <220> "Feature" and associated numeric identifiers and responses. Use of <220> to <223> is <b>MANDATORY</b> if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section. (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)	
12 <input type="checkbox"/> PatentIn 2.0 "bug"	Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.	
13 <input type="checkbox"/> Misuse of n/Xaa	"n" <b>can only represent a single nucleotide</b> ; "Xaa" <b>can only represent a single amino acid</b>	



PCT

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005  
TIME: 15:15:15

Input Set : A:\PTO.DA.txt  
Output Set: N:\CRF4\07112005\J540615.raw

4 <110> APPLICANT: CENTER FOR GENETIC ENGINEERING AND BIOTECHNOLOGY  
6 <120> TITLE OF INVENTION: RECOMBINANT HEPATITIS A VIRUS ANTIGENS PRODUCED IN PLANT  
CELLS.  
9 <130> FILE REFERENCE: ORF.  
C--> 11 <140> CURRENT APPLICATION NUMBER: US/10/540,615  
C--> 12 <141> CURRENT FILING DATE: 2005-06-23  
14 <160> NUMBER OF SEQ ID NOS: 24  
16 <170> SOFTWARE: PatentIn Ver. 2.1

## ERRORED SEQUENCES

18 <210> SEQ ID NO: 1 *pls use type*  
19 <211> LENGTH: 25  
E--> 20 <212> TYPE: ADN *DN*  
21 <213> ORGANISM: Chimeric Sequence  
24 <220> FEATURE:  
25 <221> NAME/KEY: primer\_bind  
26 <222> LOCATION: (1)..(25)  
27 <223> OTHER INFORMATION: Sequence # 1.  
28 Sequence of the oligonucleotide # 1 used for the amplification of ORF  
29 coding sequence by RT-PCR.  
32 <400> SEQUENCE: 1  
33 cttaatcttag aatgaatatg tcccaa  
36 <210> SEQ ID NO: 2  
37 <211> LENGTH: 22 *Same error*  
E--> 38 <212> TYPE: ADN *Same error*  
39 <213> ORGANISM: Chimeric Sequence *Same error*  
41 <220> FEATURE:  
42 <221> NAME/KEY: primer\_bind  
43 <222> LOCATION: (1)..(22)  
44 <223> OTHER INFORMATION: Sequence # 2.  
45 Sequence of the oligonucleotide # 2 used for the amplification of ORF  
46 coding sequence by RT-PCR.  
47 <400> SEQUENCE: 2  
48 gaaagaaaata aaggtacctc ag  
51 <210> SEQ ID NO: 3  
52 <211> LENGTH: 6685 *Same length*  
E--> 53 <212> TYPE: ADN *Same*  
54 <213> ORGANISM: Hepatitis A virus  
56 <220> FEATURE:  
57 <221> NAME/KEY: gene  
58 <222> LOCATION: Complement((1)..(6685))  
59 <223> OTHER INFORMATION: Sequence # 3.

25

22

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005  
TIME: 15:15:15

Input Set : A:\PTO.DA.txt  
Output Set: N:\CRF4\07112005\J540615.raw

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005

TIME: 15:15:15

Input Set : A:\PTO.DA.txt

Output Set: N:\CRF4\07112005\J540615.raw

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005  
TIME: 15:15:15

Input Set : A:\PTO.DA.txt  
Output Set: N:\CRF4\07112005\J540615.raw

160 acaacttgtc caaaagatga attgagacca ttagagaaa ttttggaaatc aaaaacaaga 5760  
161 gctattgtg cttgccctt ggattataca attttatgtc gaatgtattg gggccagct 5820  
162 attagttt ttcatttgaa tccagggtt cacacagggtt ttgttattgg catagatcct 5880  
163 gatagacagt gggatgaatt atttaaaaca atgataagat ttggagatgt tggcttgc 5940  
164 ttagattttt ctgctttga tgccagttt agtccattt tgatttaggaa agcaggtaga 6000  
165 atcatgagtg aattatctgg aacaccatct cattttggaa cagctttat caatactatc 6060  
166 atttattctta aacatctgtc gtacaattgt ttttattcag tctgtgggtc aatgccttct 6120  
167 gggcttcctt gtacagctt gttgaattca attattaata atattaattt gtattatgtg 6180  
168 ttttctaaaa tattttggaa gtctccagtt ttctttgtc aagctttgag gatcctttgt 6240  
169 tatggagatg atgtttgtat agtttttcc agatgttc aaattgataa tcttgacttg 6300  
170 attggacaga aaatttggaa tgagttcaaa aaacttggca tgacagccac tttagctgac 6360  
171 aaaaatgtgc ctcaactgaa gccagttca gaattgactt ttcttaaaag atcttttaat 6420  
172 ttgggtggagg acagaatcag acctgcaatt tcagaaaaga caatttggc tttgatagct 6480  
173 tggcagagaa gtaacgctga gtttggcag aatttagaaa atgctcagtg gtttgccttc 6540  
174 atgcatggct atgagttcta tcagaaattc tatttttg ttcaagtctg tttggagaaaa 6600  
175 gagatgatag aatatagact taaatctt gattttggaa gaatgagatt ttatgaccag 6660  
176 tgtttcattt gtgacccccc atgt 6685

pls use type DNA

E--> 181 <212> TYPE: ADN

182 <213> ORGANISM: Chimeric Sequence

same error

184 <220> FEATURE:

185 <221> NAME/KEY: primer\_bind

186 <222> LOCATION: (1)..(40)

187 <223> OTHER INFORMATION: Sequence # 4.

188 Sequence of the oligonucleotide # 5 used for the  
189 amplification of P1-2A coding sequence by PCR.

192 <400> SEQUENCE: 4

193 ttgaattcag cttgtgaaaa taacccttc atttccttag

40

196 <210> SEQ ID NO: 5

197 <211> LENGTH: 28

Same error

E--> 198 <212> TYPE: ADN

199 <213> ORGANISM: Chimeric Sequence

201 <220> FEATURE:

202 <221> NAME/KEY: primer\_bind

203 <222> LOCATION: (1)..(28)

204 <223> OTHER INFORMATION: Sequence # 5.

206 Sequence of the oligonucleotide # 5 used for the  
207 amplification of P1-2A coding sequence by PCR.

209 <400> SEQUENCE: 5

210 cggccgggtc tagaatgaat atgtccaa

28

213 <210> SEQ ID NO: 6

214 <211> LENGTH: 2523

Same error

E--> 215 <212> TYPE: ADN

216 <213> ORGANISM: Hepatitis A virus

218 <220> FEATURE:

219 <221> NAME/KEY: gene

220 <222> LOCATION: Complement((1)..(2523))

221 <223> OTHER INFORMATION: Sequence # 6.

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005  
TIME: 15:15:15

Input Set : A:\PTO.DA.txt  
Output Set: N:\CRF4\07112005\J540615.raw

222 Nucleotide sequence coding for the structural  
223 P1-2A HAV proteins of the M2 strain.  
226 <400> SEQUENCE: 6  
227 atgaatatgt ccaaacaagg aattttccag actgttggga gtggcctgta ccacatcctg 60  
228 tccttggcag atattgagga agagcaaatg attcagtccg ttgataggac tgcagtgact 120  
229 ggagcttctt atttcacttc tgtggaccaa tcttcagttc atactgctgta ggttggctca 180  
230 caccaaattt aacccttgaa aacctctgtt gataaaacctg gttctaagaa aactcagggg 240  
231 gagaagttt tcttgattca ttctgctgat tggctacta cacatgctct ctttcatgaa 300  
232 gttgcaaaat tggatgtgg tggactgctg tacaatgagc agtttgcgtt ccaagggttt 360  
233 tttagatacc atacttatgc aagatttggc atttagatcc aagttcagat aaatcccaca 420  
234 cccttcagc aaggaggact aatctgtgcc atggccctg gtgaccaaag ttatggttca 480  
235 atagcatcct tgactgtta tcctcatggt ctgttaattt gcaatataaa caatgttagtt 540  
236 agaataaaagg ttccattttt ttatactaga ggtgcttac attttaaaga tccacagttac 600  
237 ccagtttggg aattgacaat cagagtttgg tcagagttga atattggAAC aggaacctca 660  
238 gtttataactt cactcaatgtt ttttagcttggg tttacagatt tggagtttca tggatttaact 720  
239 cctctttctt cacagatgtt gagaatgaa ttttagatgg tttactactga aaatgttggta 780  
240 aatttgcattt attatgaaga tgcaggggca aaaatgttctt ttgctttggta tcaggaagat 840  
241 tggaaagtctg atccctccca aggtgggttga attaaaattt ctcatcatttac tacctggaca 900  
242 tccattccaa ccttagctgc tcagtttcca ttcaatgctt cagattcaatg tggcaaccaa 960  
243 attaaagtta taccagtggc cccatactttt ttccagatgta caaaacactaa tcctgatcaa 1020  
244 aaatgtataa cagccttggc ctctattttt cagatgttctt gttttggag gggagatctt 1080  
245 gttttcgatt tccaggtttt tccaaacaaa tatcatttcag gttagctgtt gttttgtttt 1140  
246 gttccttgggaa atgagttat agatgttact ggaattacat taaaacaggc aactactgct 1200  
247 ccttgtgcag tggatggacat tacaggagtg cagtcacactt tgagatttcg tggatggat 1260  
248 atttctgata caccctatcg agtgaatagg tacacgaatg cagcacatca aaaaggtgag 1320  
249 tatactgcca ttggggaaagct tattgtgtat tggttataata gattgacttc tccttcta 1380  
250 gttgcttctc atgtagttagt taatgtttt ctttcagca ttaatttggta atggtttgtt 1440  
251 cctctttacc atgctatggc tggttaccaca caggttggag atgatttcagg aggtttctca 1500  
252 acaacagttt ctacagagca gaatgttctt gatccccaaatg tggcataac aaccatgagg 1560  
253 gattttaaag ggaaagccaa tagggggaaag atggatgtat caggagtgca ggtacctgtg 1620  
254 ggagcttatttta caacaatttgc ggttccaggat tttagcaaaaga aagtacatgacatcc 1680  
255 gaatttgcattt ctggagaatc cagacatataca tcagatcaca tgtctatttta taaatttcatg 1740  
256 ggaaggcttc atttcttgcg tacttttact tttaatttcaaa acaataaaga gtacacattt 1800  
257 ccaataactc tgtcttcgac ttctaaatcctt cctcatgggtt taccatcaac attaagggtgg 1860  
258 ttctttaattt tggttcaatgtt gtatagagga ccattggatt tgacaattt aatcacagga 1920  
259 gccactgatg tggatgttat ggcctgggtt actccagtgg gccttgcgtt cgacacccct 1980  
260 tgggtggaaa agaagtccgc ttgttctattt gattataaaa ctgccttgg agctgttaga 2040  
261 tttaatacaa gaagaacagg gaacatttcatttgc catggatttc ttattttgtat 2100  
262 ggcgtgtctg gggacttggc tggcttggga gataagacatg attctacattt tggattggat 2160  
263 tctatttcaga ttgcaaaatttta caatcatttcatttgc gatgtatattt tgccctttag ttgttattttg 2220  
264 ttgttcacatc agcaatcaga gttctatttcatttgc ctttagatcatttgc cattaaatttcaaaatgtatg 2280  
265 ttgtccactg agtccatgtat ggttgcatttgc ggttgcgtt ggttgcgtt ggttgcgtt 2340  
266 gatcccagat cagaggagga cagaagatattt gggatgtat ggttgcgtt ggttgcgtt 2400  
267 aaagaatttgc gactggatgtt gggaaacaa agaattcaat atgctcaggat ggttgcgtt 2460  
268 aatgaagtgc ttccaccccttcc tagggaaatg aagggttgc ggttgcgtt ggttgcgtt 2520  
269 gat  
272 <210> SEQ ID NO: 7  
273 <211> LENGTH: 27  
E--> 274 <212> TYPE: ADN *Same error*

## VERIFICATION SUMMARY

PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005

TIME: 15:15:16

Input Set : A:\PTO.DA.txt  
Output Set: N:\CRF4\07112005\J540615.raw

L:11 M:270 C: Current Application Number differs, Replaced Application Number  
L:12 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:20 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:38 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:53 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:181 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:198 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:215 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:274 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:291 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:308 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:328 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:348 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:365 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:382 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:416 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:492 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:508 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:525 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:535 M:259 W: Allowed number of lines exceeded, <223> Other Information:  
L:600 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:617 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:633 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:637 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:20  
L:650 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:654 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:21  
L:667 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:726 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:  
L:759 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005  
TIME: 15:15:15

Input Set : A:\PTO.DA.txt  
Output Set: N:\CRF4\07112005\J540615.raw

632 <211> LENGTH: 55 ←  
E--> 633 <212> TYPE: ADN ← Same err.↓  
634 <213> ORGANISM: Chimeric Sequence  
636 <220> FEATURE:  
W--> 637 <221> NAME/KEY: D\_segment  
638 <222> LOCATION: (1)..(54)  
639 <223> OTHER INFORMATION: Sequence # 20  
640 synthetic fragment modifying the 3' end of  
641 the 2A protein and introduces a space-bar  
642 between this one and the KDEL signal.  
644 <400> SEQUENCE: 20  
645 ccttagaaaaa tgaagggtt atatgcttct ggaggtgaat tcgatatcaa ggatg 55

The type of errors shown exist throughout  
the Sequence Listing. Please check subsequent  
sequences for similar errors.

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005  
TIME: 15:15:15

Input Set : A:\PTO.DA.txt  
Output Set: N:\CRF4\07112005\J540615.raw

573 attcagattt caaattacaa tcattctgtat gaatattttgt ccttttagttt ttatgggtct 2160  
 574 gtcacagagc aatcagagtt ctatccctt agagctccat taaattcaaa tgctatgtt 2220  
 575 tccactgagt ccatgtatgag tagaattgca gctggagact tggagtcatc agtggatgt 2280  
 576 cccagatcg aggaggacag aagattttagt agtcatatag aatgttagaa accatataaa 2340  
 577 gaatttgagac tggaggttgg gaaacaaaga atcaaataatg ctcaggaaga gttatcaa 2400  
 578 gaagtgcctt caccccttag gaaaatgaag ggttatttt cacaagctga attccctgcag 2460  
 579 cccggggat ccatgggaat ttcatggat gacaatgata gtgcagttagc tgagttttc 2520  
 580 cggcttttc catctggta accatcaaattt tccaaaggat ttcttggatt gctatgg 2580  
 581 actaatcaca agtgggttgc tgtggagct gcagttggta aaccaattcc agtgttgg 2640  
 582 ggtatggttt tggatggatgg ttttcccgaa aagaggaaatg aaccaattcc agtgttgg 2700  
 583 gtttatcatg ggtgtactaa gcccaaaccat gttgatggat tggatgcaga tccagtagag 2760  
 584 tctcgttgc ctctagaaat agcaggatata gtttagaaaa atttggatc gtttggagtt 2820  
 585 ggtgagaaaaa atggatgtgtt gggatggatc atgaatgcct taggagtgaa ggatgattgg 2880  
 586 ttgttagtac cttctcatgc ttataaattt gaaaaggattt atgaaatgtt gggatggat 2940  
 587 ttcaatagag gtggaaacttta ttatcaattt tcagttggta atgttggat tcaatctta 3000  
 588 gatgtggat tccaaagatgt tggatggatc aaggttccat caattccaa gtttagagat 3060  
 589 attactcaac attttattaa gaaaggagat gtgccttagat ctttgaatcg ctggcaaca 3120  
 590 ttagtgacaa ccgttaatgg aactcctatg ttaatttctg agggacctt aaaaatggaa 3180  
 591 gaaaaagcca cttatgttca taagaagaat gatggacta cgggtgattt gactgttagat 3240  
 592 caggcatgga gaggaaaaagg tgaaggtttt cctggatgtt gttgggggc cctagtgtca 3300  
 593 tcaaatcagt ccataaaaaa tgcaattttt ggttattcatg ttgctggagg aaattcaattt 3360  
 594 cttgtggcaa agttgatttac tcaagaaatg tttcaaaaaca ttgataagaa aattgaaatc 3420  
 595 aagctt  
 596 <210> SEQ ID NO: 18  
 597 <211> LENGTH: 19  
 E--> 600 <212> TYPE: ADN  
 601 <213> ORGANISM: Chimeric Sequence  
 603 <220> FEATURE:  
 604 <221> NAME/KEY: sig\_peptide  
 605 <222> LOCATION: (1)..(19)  
 606 <223> OTHER INFORMATION: Sequence #18.  
 607 Synthetic fragment corresponding to the  
 608 KDEL endoplasmic reticulum retention signal  
 609 sequence.  
 610 <400> SEQUENCE: 18  
 611 atcaaggatg aattgtat 19  
 612 atcaaggatg aattgtat  
 613 <210> SEQ ID NO: 19  
 614 <211> LENGTH: 21  
 E--> 615 <212> TYPE: ADN  
 616 <213> ORGANISM: Chimeric Sequence  
 618 <220> FEATURE:  
 619 <221> NAME/KEY: sig\_peptide  
 620 <222> LOCATION: (1)..(21)  
 621 <223> OTHER INFORMATION: Sequence #19.  
 622 Synthetic fragment corresponding to the KDEL  
 623 endoplasmic reticulum retention signal sequence.  
 624 <400> SEQUENCE: 19  
 625 cgattacaat tcatccttga t  
 631 <210> SEQ ID NO: 20

19

21

**RAW SEQUENCE LISTING**  
**PATENT APPLICATION: US/10/540,615**

DATE: 07/11/2005  
TIME: 15:15:15

Input Set : A:\PTO.DA.txt  
Output Set: N:\CRF4\07112005\J540615.raw

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005  
TIME: 15:15:15

Input Set : A:\PTO.DA.txt  
Output Set: N:\CRF4\07112005\J540615.raw

462 tgggtggaaa agaagtcagc tttgtctatt gattataaaa ctgcccttgg agctgttaga 2040  
463 ttaataacaa gaagaacagg gaacattcag attagattgc catggatttc ttatttgtat 2100  
464 gccgtgtctg gagcaacttgg a tggcttggg gataagacag attctacatt tggattggtt 2160  
465 tctattcaga ttgcaaatta caatcattct gatgaatatt tgcctttag ttgttatttg 2220  
466 tctgtcacag agcaatcaga gttctatttc cctagagctc cattaaattc aaatgctatg 2280  
467 ttgtccactg agtccatgtat gatgagaatt gcagctggag acttggagtc atcagtggat 2340  
468 gatcccagat cagaggagga cagaagattt gagagtcata tagaatgtag gaaaccatata 2400  
469 aaagaattga gactggaggt tggaaacaa agaatcaa at gctcagga agagttatca 2460  
470 aatgaagtgc ttccacctcc taggaaaatg aagggttat tttcacaagc tgaattcctg 2520  
471 cagcccgaaa gatccatggg aatttcagat gatgacaatg atagtgcagt agtgcgtt 2580  
472 ttccggctt ttccatctgg tgaaccatca aattccaatg tatctagtt tttccaagct 2640  
473 gtcactaattc acaagtgggt tgctgtggg gctgcagttt gtattcttgg attgcttagt 2700  
474 ggaggatggt ttgtgtataa gcattttcc cgcaaagagg aagaaccaat tccagctgtt 2760  
475 gggtttatac atggagtgtac taagccaaa caagtgatta aattggatgc agatccagta 2820  
476 gagtctcagt tgactctaga aatagcagga tttagtttaga aaaattttgt tcagtttgg 2880  
477 ttgggtgaga aaaatggatg tggatgtt gtcatgaatg ctttagggat gaaggatgtt 2940  
478 ttgggtttag taccttctca tgcttataaa tttgaaaagg attatgaaat gatggatgtt 3000  
479 tatttcaata gaggtggaaac ttactattca atttcagttt gtaatgttgt tattcaatct 3060  
480 ttagatgtgg gattccaaga tgggtttctt atgaagggtt ctacaatcc caagttttaga 3120  
481 gatattactc aacattttat taagaaagga gatgtgccta gggccttggaa tcgcttggca 3180  
482 acattagtga caaccgttaa tggaaactcct atgtaattt ctgagggacc tttaaaaatg 3240  
483 gaagaaaaag ccacttatgt tcataagaag aatgatggta ctacgggttga tttgactgt 3300  
484 gatcaggcat ggagaggaaa aggtgaaggt cttcttggaa tggatgttgg ggccttagt 3360  
485 tcatcaaattc agtccatataca aatgcaatt ttgggttattc atgatgttgg agggaaattca 3420  
486 attcttggaa caaagttgtat tactcaagaa atgatgttgg acattgataa gaaaatttggaa 3480  
487 atcaagctt 3489

490 <210> SEQ ID NO: 15

491 <211> LENGTH: 51

← Same errors

E--> 492 <212> TYPE: ADN

493 <213> ORGANISM: Chimeric Sequence

495 <220> FEATURE:

496 <221> NAME/KEY: gene

497 <222> LOCATION: (1)..(51)

498 <223> OTHER INFORMATION: Sequence # 15.

499 Synthetic fragment that reverts the

500 transcription start of the vp2 protein.

502 <400> SEQUENCE: 15

503 gggatggata ttgaggaaga gcaaatttattt cagtcgttgc ataggactgc a

51

506 <210> SEQ ID NO: 16

507 <211> LENGTH: 47

←

Same errors

E--> 508 <212> TYPE: ADN

509 <213> ORGANISM: Chimeric Sequence

511 <220> FEATURE:

512 <221> NAME/KEY: gene

513 <222> LOCATION: (1)..(47)

514 <223> OTHER INFORMATION: Sequence # 16.

515 Synthetic fragment that reverts the transcription

516 start of the vp2 protein (complementary chain).

519 <400> SEQUENCE: 16

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005  
TIME: 15:15:15

Input Set : A:\PTO.DA.txt  
Output Set: N:\CRF4\07112005\J540615.raw

✓ Same  
error

346 <210> SEQ ID NO: 11  
347 <211> LENGTH: 25  
E--> 348 <212> TYPE: ADN  
349 <213> ORGANISM: Chimeric Sequence  
351 <220> FEATURE:  
352 <221> NAME/KEY: primer\_bind  
353 <222> LOCATION: (1)..(25)  
354 <223> OTHER INFORMATION: Sequence # 11.  
355 Sequence of the oligonucleotide # 11 used  
356 for the amplification of 3C coding sequence by PCR  
359 <400> SEQUENCE: 11  
360 tctcagtcaa ctctagaaat agcag 25  
363 <210> SEQ ID NO: 12  
364 <211> LENGTH: 21  
E--> 365 <212> TYPE: ADN  
366 <213> ORGANISM: Chimeric Sequence  
368 <220> FEATURE:  
369 <221> NAME/KEY: primer\_bind  
370 <222> LOCATION: (1)..(21)  
371 <223> OTHER INFORMATION: Sequence # 12.  
372 Sequence of the oligonucleotide # 12 used for  
373 the amplification of 3C coding sequence by PCR  
376 <400> SEQUENCE: 12  
377 ataagcttga tcaattttct t 21  
380 <210> SEQ ID NO: 13  
381 <211> LENGTH: 978  
E--> 382 <212> TYPE: ADN  
383 <213> ORGANISM: Hepatitis A virus  
385 <220> FEATURE:  
386 <221> NAME/KEY: gene  
387 <222> LOCATION: Complement((1)..(978))  
388 <223> OTHER INFORMATION: Sequence # 13.  
389 Sequence corresponding to the region of 3ABC  
390 polyprotein with proteolytic activity having  
391 the selfprocessing sites mutated.  
394 <400> SEQUENCE: 13  
395 gaattcctgc agccccgggg atccatggga atttcagatg atgacaatga tagtgcagta 60  
396 gctgagttt tccggcttt tccatctggt gaaccatcaa attccaagtt atctagttt 120  
397 ttccaagctg tcactaatca caagtgggtt gctgtggag ctgcagttgg tattcttgg 180  
398 ttgctagtgg gaggatgggtt tgtgtataag cattttccc gcaaagagga agaaccaatt 240  
399 ccagctgttg gggtttatca tggagtgact aagcccaaac aagtgataa attggatgca 300  
400 gatccagtag agtctcagtt gactctagaa atagcaggat tagtttagaa aaatttgggtt 360  
401 cagtttggag ttgggtgagaa aaatggatgt gtgagatggg tcatgaatgc cttaggagt 420  
402 aaggatgatt ggttggtagt accttctcat gcttataat ttgaaaagga ttatgaaatg 480  
403 atggagttt attcaatag aggtggaaact tactattcaa tttcagctgg taatgttgg 540  
404 attcaatctt tagatgtggg attccaagat gttgttctaa tgaagggtcc tacaattccc 600  
405 aagtttagag atattactca acattttatt aagaaaaggag atgtgcctag agccttgaat 660  
406 cgcttggcaa cattagtgac aaccgttaat ggaactccta tgtaatttc tgagggacct 720  
407 taaaaatgg aagaaaaagc cacttatgtt cataagaaga atgatggatc tacgggttcat 780

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005  
TIME: 15:15:15

Input Set : A:\PTO.DA.txt  
Output Set: N:\CRF4\07112005\J540615.raw

408 ttgactgttag atcaggcatg gagagggaaaa ggtgaaggc ttcctggaat gtgtgggtggg 840  
 409 gccctagtgt catcaaatca gtcatacaa aatgcaattt tgggtattca tggtgctgga 900  
 410 ggaaattcaa ttcttgccaa aaagttgatt actcaagaaa tggttcaaaa cattgataag 960  
 411 aaaattgaaa tcaagctt 978  
 414 <210> SEQ ID NO: 14  
 415 <211> LENGTH: 3489  
 E--> 416 <212> TYPE: ADN *same error*  
 417 <213> ORGANISM: Hepatitis A virus  
 419 <220> FEATURE:  
 420 <221> NAME/KEY: gene  
 421 <222> LOCATION: Complement((1)..(3489))  
 422 <223> OTHER INFORMATION: Sequence # 14.  
 423 Nucleotide sequence CODING for the new  
 424 modified open reading frame (ORFm) of the  
 425 Cuban M2 strain.  
 428 <400> SEQUENCE: 14  
 429 atgaatatgt ccaaacaagg aattttccag actgttggga gtggccttga ccacatcctg 60  
 430 tccttggcag atattgagga agagcaaatg attcagtccg ttgataggac tgcagtgact 120  
 431 ggagcttctt atttcacttc tggacccaa tcttcagttt atactgctga gttggctca 180  
 432 caccaaattt aacctttgaa aacctctgtt gataaacctg gttctaaagaa aactcagggg 240  
 433 gagaagttt tcttgatca ttctgctgtt tggctcaacta cacatgctct ctttcatgaa 300  
 434 gttgaaaat tggatgtgtt gaaactgtgt tacaatgagc agtttgcgtt ccaagggttt 360  
 435 ttgagatacc atacttatgc aagatttggc attgagattc aagttcagat aaatcccaca 420  
 436 cccttcagc aaggaggact aatctgtgcc atggttcctg gtgaccaaaat ttatggttca 480  
 437 atagcatcct tgactgttt tcctcatgtt ctgttaaattt gcaatataaa caatgttagtt 540  
 438 agaataaagg ttccattttt ttatactaga ggtgcttatac attttaaaga tccacagtac 600  
 439 ccagttggg aattgacaat cagagttggg tcagagttga atattggaaac aggaacctca 660  
 440 gcttatactt cactcaatgt tttagctagg tttacagatt tggagttgca tggattaact 720  
 441 cctcttcttca cacagatgtt gggaaatgaa ttttagagttt gtactactga aaatgtgtt 780  
 442 aatttgtcaa attatgaaga tgcaggggca aaaatgtctt ttgctttggaa tcaggaagat 840  
 443 tggaaagtctg atccttccca aggtgggtggaa attaaaattt ctcatttcac tacctggaca 900  
 444 tccattccaa ccttagctgc tcagtttcca ttcaatgtttt cagattcagt tgggcaacaa 960  
 445 attaaagtta taccagtggc cccatacttt ttccagatga caaacactaa tcctgatcaa 1020  
 446 aaatgtataa cagccttggc ctctattttgt cagatgttctt gctttggag gggagatctt 1080  
 447 gtttgcatt tccaggtttt tccaaacaaa ttcatttcag gtaggctgtt gttttgtttt 1140  
 448 gttcctggaa atgagttat agatgttact ggaattacat taaaacagggc aactactgt 1200  
 449 ctttgtcag tggatggacat tacaggagtg cagtcaacct tgagattcg tggctttgg 1260  
 450 atttctgata caccctatcg agtgaatagg tacacgaagt cagcacatca aaaaggtgag 1320  
 451 tatactgcca ttggaaagct tattgtgtat ttttataata gattgacttc tccttctaatt 1380  
 452 gttgcttctc atgtagttagt taatgtttat ctttcagca ttaatttggaa atggtttgtt 1440  
 453 cctcttacc atgtatggc tggatggacat caggtggag atgattcagg aggtttctca 1500  
 454 acaacagttt ctacagagca gaatgttccct gatccccaaat ttggcataac aaccatgagg 1560  
 455 gattaaaag ggaaagccaa tagggggaaat atggatgtat caggagtgc ggtacctgtg 1620  
 456 ggagcttata caacaattga ggatccagg ttagcaaaga aagtacactga gacatttccct 1680  
 457 gaattgaagc ctggagaatc cagacataca tcagatcaca tggatgttta taaatttcatg 1740  
 458 ggaaggtctc atttcttgc tttttttact ttaatttcaa acaataaaga gtacacattt 1800  
 459 ccaataactc tggatggc tttcaatccct cttcatgtt taccatcaac attaaggtgg 1860  
 460 ttctttaatt tggatgttgc ttatagagggaa ccattggatt tgacaattat aatcacaggg 1920  
 461 qccactqatq tggatgttgc ttatagagggaa actccatqg qccctqctgt cgacacccct 1980

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/540,615

DATE: 07/11/2005

TIME: 15:15:15

Input Set : A:\PTO.DA.txt

Output Set: N:\CRF4\07112005\J540615.raw

275 <213> ORGANISM: Chimeric Sequence ← Same error.

277 <220> FEATURE:

278 <221> NAME/KEY: primer\_bind

279 <222> LOCATION: (1)..(27)

280 <223> OTHER INFORMATION: Sequence # 7.

281 Sequence of the oligonucleotide # 7 used for

282 the amplification of 3A coding sequence by PCR.

285 <400> SEQUENCE: 7

286 ccatggaaat ttcagatgtat gacaatg 27

289 <210> SEQ ID NO: 8

290 <211> LENGTH: 26 ← Same error

E--> 291 <212> TYPE: ADN

292 <213> ORGANISM: Chimeric Sequence

294 <220> FEATURE:

295 <221> NAME/KEY: primer\_bind

296 <222> LOCATION: (1)..(26)

297 <223> OTHER INFORMATION: Sequence # 8.

298 Sequence of the oligonucleotide # 7 used for

299 the amplification of 3A coding sequence by PCR.

302 <400> SEQUENCE: 8

303 ggatatatcggt tcttctcttt tgccggg 26

306 <210> SEQ ID NO: 9

307 <211> LENGTH: 85 ← Same error

E--> 308 <212> TYPE: ADN

309 <213> ORGANISM: Chimeric Sequence

312 <220> FEATURE:

313 <221> NAME/KEY: gene

314 <222> LOCATION: (1)..(85)

315 <223> OTHER INFORMATION: Sequence # 9.

316 Synthetic fragment coding for 3B protein

317 carrying T by C and G by C nucleotide

318 substitutions, respectively.

321 <400> SEQUENCE: 9

322 tccagctgtt ggggtttatc atggagtgtac taagccaaa caagtgattt aattggatgc 60

323 agatccatgtt gacttcgtt tgact 85

326 <210> SEQ ID NO: 10

327 <211> LENGTH: 89 ← Same error

E--> 328 <212> TYPE: ADN

329 <213> ORGANISM: Chimeric Sequence

332 <220> FEATURE:

333 <221> NAME/KEY: gene

334 <222> LOCATION: (1)..(89)

335 <223> OTHER INFORMATION: Sequence # 10.

336 Synthetic fragment coding for 3B protein

337 carrying T by C and G by C nucleotide

338 substitutions, respectively (complementary chain).

341 <400> SEQUENCE: 10

342 ctagagtcaa ctgagactct actggatctg catccaattt aatcaacttgt ttgggcttag 60

343 tcactccatgtt ataaacccca acagctgg 89